Published online 31 May 2012 *Nucleic Acids Research, 2012, Vol. 40, Web Server issue* ***W597–W603***
doi:10.1093/nar/gks400

ExPASy: SIB bioinformatics resource portal

Panu Artimo[1], Manohar Jonnalagedda[1,2], Konstantin Arnold[3], Delphine Baratin[4], Gabor Csardi[5], Edouard de Castro[4], Séverine Duvaud[4], Volker Flegel[1], Arnaud Fortier[1], Elisabeth Gasteiger[4], Aurélien Grosdidier[2], Céline Hernandez[1], Vassilios Ioannidis[1], Dmitry Kuznetsov[1], Robin Liechti[1], Sébastien Moretti[1,6], Khaled Mostaguir[4], Nicole Redaschi[4], Grégoire Rossier[1], Ioannis Xenarios[1,4,7] and Heinz Stockinger[1,*]

[1]Vital-IT Group, SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland, [2]Molecular Modelling Group, SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland, [3]Computational Structural Biology Group, University of Basel, SIB Swiss Institute of Bioinformatics, Basel, Switzerland, [4]Swiss-Prot Group, SIB Swiss Institute of Bioinformatics, Geneva, Switzerland, [5]Department of Statistics, Harvard University, Cambride, USA, [6]Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland and [7]Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland

Received December 16, 2011; Revised April 10, 2012; Accepted April 17, 2012

ABSTRACT

ExPASy (http://www.expasy.org) has worldwide reputation as one of the main bioinformatics resources for proteomics. It has now evolved, becoming an extensible and integrative portal accessing many scientific resources, databases and software tools in different areas of life sciences. Scientists can henceforth access seamlessly a wide range of resources in many different domains, such as proteomics, genomics, phylogeny/evolution, systems biology, population genetics, transcriptomics, etc. The individual resources (databases, web-based and downloadable software tools) are hosted in a 'decentralized' way by different groups of the SIB Swiss Institute of Bioinformatics and partner institutions. Specifically, a single web portal provides a common entry point to a wide range of resources developed and operated by different SIB groups and external institutions. The portal features a search function across 'selected' resources. Additionally, the availability and usage of resources are monitored. The portal is aimed for both expert users and people who are not familiar with a specific domain in life sciences. The new web interface provides, in particular, visual guidance for newcomers to ExPASy.

INTRODUCTION

Historically, ExPASy (1,2) was one of the first web servers in the life sciences community, known as 'Expert Protein Analysis System'. We have now turned it into a bioinformatics resource portal that serves not only the field of proteomics but also other domains of life sciences. The major changes and their rationale are described below:

(i) comprehensive list of high-quality resources provided by the SIB Swiss Institute of Bioinformatics and several external institutions. Swiss-Prot has been and still is the flagship knowledge base of the SIB and one of the main proteomics-related resources on ExPASy. However, in the last decade, the SIB has grown to a Swiss-wide institute that covers all major bioinformatics domains, including proteomics, genomics, transcriptomics, evolution, systems biology, etc. One goal of the new ExPASy portal is to serve as entry point to all scientific databases and tools provided by the SIB, and we revisited the original focus on proteomics accordingly;

(ii) federated portal rather than a single server. The original ExPASy server hosted resources from mainly two groups located in Geneva (with a few notable exceptions, such as SWISS-MODEL). With the inclusion of resources from more than 20 groups in five different cities (Geneva, Lausanne,

*To whom correspondence should be addressed. Tel: +41 21 692 40 89; Fax: +41 21 692 40 65. Email: Heinz.Stockinger@isb-sib.ch; helpdesk@expasy.org

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

Berne, Basel and Zurich), located at seven Swiss higher education institutions, ExPASy has been turned into a federated portal that provides access to resources at different locations. Each resource is maintained independently but many of them still follow a common release cycle or database update schedule. Keeping independent release cycles for some resources allows more flexibility for the individual resource providers;

(iii) new search features: a flexible search functionality is provided to allow proper access to the independent resources now federated under the new ExPASy portal. This relies on a standardized search interface that permits querying several different resources in parallel;

(iv) visual guidance: we aim at serving a wide range of users including newcomers not familiar with the fields covered by ExPASy. We have added an attractive visual interface that is addressed to new users even if they have little experience with the provided domains and technologies; and

(v) monitor the quality and usage of resources: providing high-quality usable resources is of major importance. Consequently, we regularly monitor all resources on the portal and provide the possibility to notify the resource providers in case of availability issues. Additionally, usage statistics are collected to monitor the popularity of resources.

Figure 1 shows the homepage of the new version of ExPASy, including a new logo reflecting the SIB's corporate identity. Different 'categories' beyond proteomics can be selected in the menu item 'Categories' on the left-hand side (cf. 'Access to a large set of bioinformatics resources' section). A query feature is directly accessible from the page header (cf. 'Cross-resource search interface' section). In the centre of the page, a set of resources is promoted ('Featuring today') to allow users to discover new resources, which might help them in their work. These featured links are updated every week and draws from the list of all SIB resources available via the portal. Additionally, the most 'Popular resources' are directly accessible in a pane on the right side. An alphabetical resources list can be accessed under the menu item 'Resources A..Z'. Finally, 'Latest News' about resources are displayed and made available via an RSS feed (see RSS icon). In the remainder of this article, we will present details of the new features as well as some hints for experienced users of the previous ExPASy version.

MAIN FEATURES

Access to a large set of bioinformatics resources

The previous ExPASy server was mainly dedicated to bioinformatics related to proteomics (protein identification, post-translational modifications, etc.), and only

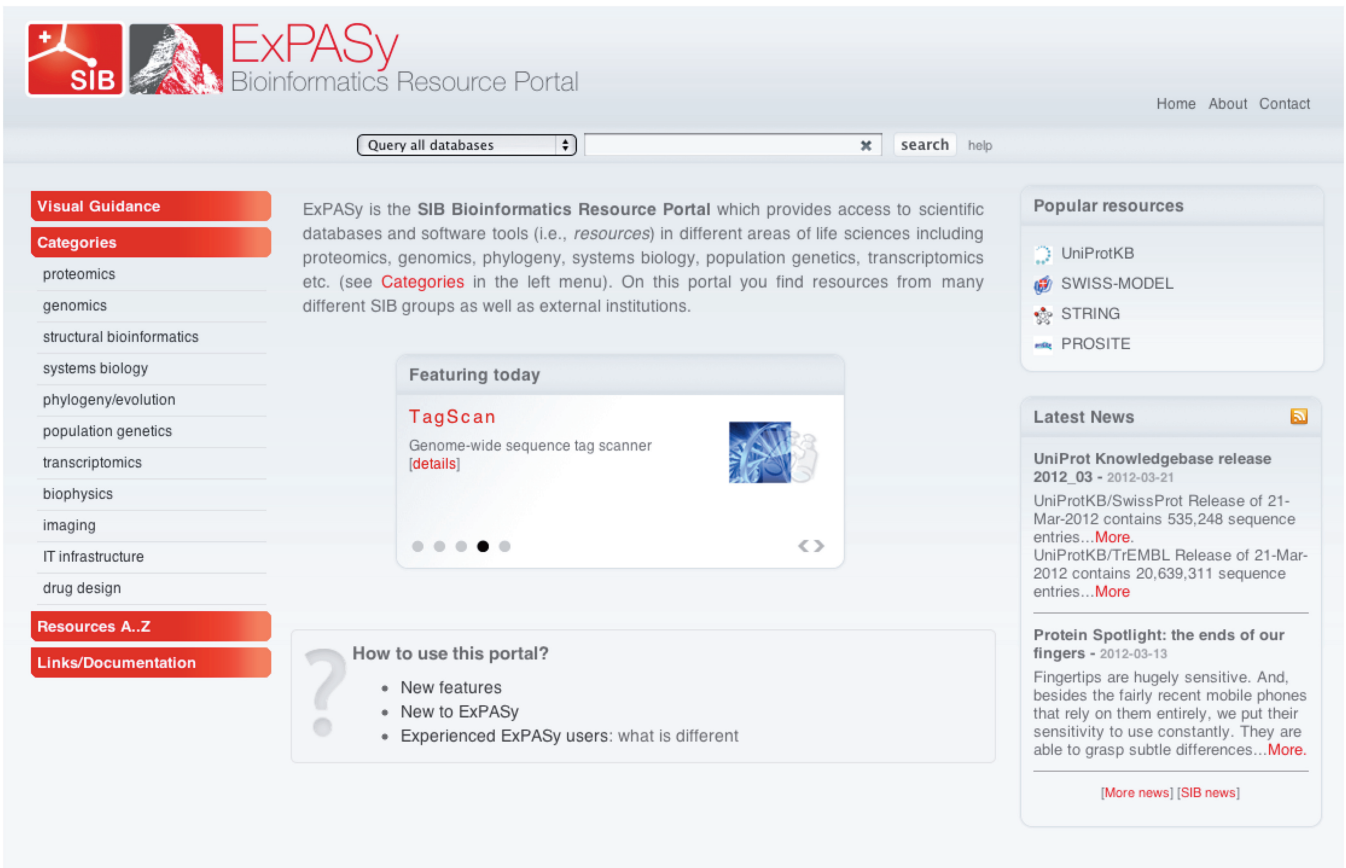

Figure 1. ExPASy homepage.

some of the resources were applicable to both proteomes and genomes (like sequence similarity search or sequence alignment). The new portal provides access to over 130 resources (cf. Appendix A1 for the full list of resources) from more than 20 different SIB groups. To reflect the different scientific directions of the SIB's groups, the following scientific categories are currently available: proteomics, genomics, structural bioinformatics, systems biology, phylogeny/evolution, population genetics, transcriptomics, biophysics, imaging and drug design. To facilitate access to the numerous resources in proteomics and genomics, additional subcategories are available in the respective menu items.

Within each category, resources are classified as 'Databases' and/or as 'Tools'. Most resources have web interfaces (especially databases) but others may consist of a downloadable software package with a command line or graphical interface.

Each single resource is documented with the following information to allow for a more specific identification and description (Figure 2).

(i) Resource name and description.
(ii) Indication of SIB group that maintains the resource.
(iii) Scientific category.
(iv) Keywords: each resource is tagged with terms from a controlled vocabulary to facilitate resource classification and search beyond simple categorization e.g. sequence comparison, ChIP-seq, etc.
(v) URL for the web interface and for download if available.
(vi) Software type: a resource can either be a website, command line interface, GUI, library, etc.
(vii) Status (green check box) information to check if a resources is accessible [cf. 'Monitoring (availability and usage checks)' section].

A detailed search feature to find resources on the portal is available in the page header (see 'search' button in top/centre of the ExPASy home page). One needs to select 'Find Resources' to discover (query) resources by name, keyword, category or description. If this feature is used, the information shown below becomes visible (Figure 2). If a search term is found (for instance 'UniProtKB'), it is highlighted for the specific resource(s). The second main usage of the query engine will be described in the 'Cross-resource search interface' section.

Internal implementation detail: all the resource-specific information is stored in a relational database back-end. This allows for a very dynamic website in case any of the 130 or more resources change. The portal is implemented using the web framework CakePHP (http://cakephp.org).

Cross-resource search interface

Although most of the resources provide specific search functionalities, a single search querying a set of resources at the same time is often more convenient. For instance, the search can even include resources that might not yet be known to the user but still be useful. Following this idea, a REST-based cross-resource search protocol (The detailed specification of the interface can be found at: http://wiki.isb-sib.ch/web-team/Sib-resource_query-interface) (Figure 3) has been designed. It is already implemented by almost 20 SIB resources, such as ENZYME, MyHits, STRING, UniProtKB, ViralZone, PROSITE, SWISS-MODEL Repository, etc.

This query feature is available in the header of the page via the item 'Query all databases' (Figure 3). In this case, a standard text-based query e.g. 'human' is sent in parallel to all compatible resources, i.e. the ones that implement the ExPASy query interface. As soon as one of the latter has performed the search, the number of matches (hits) is displayed, and a link is added to the query results in the resource's original interface. If available, an additional description of the result set is shown. The query can also be restricted by specifying a category, e.g. only query resources in the field of 'phylogeny/evolution' or even by specifying an individual resource.

In addition to the default behaviour (searching for full-text matches), the search engine automatically recognizes formatted data types (called 'query types'), such as UniProtKB accession numbers, PDBID, EnsemblID,

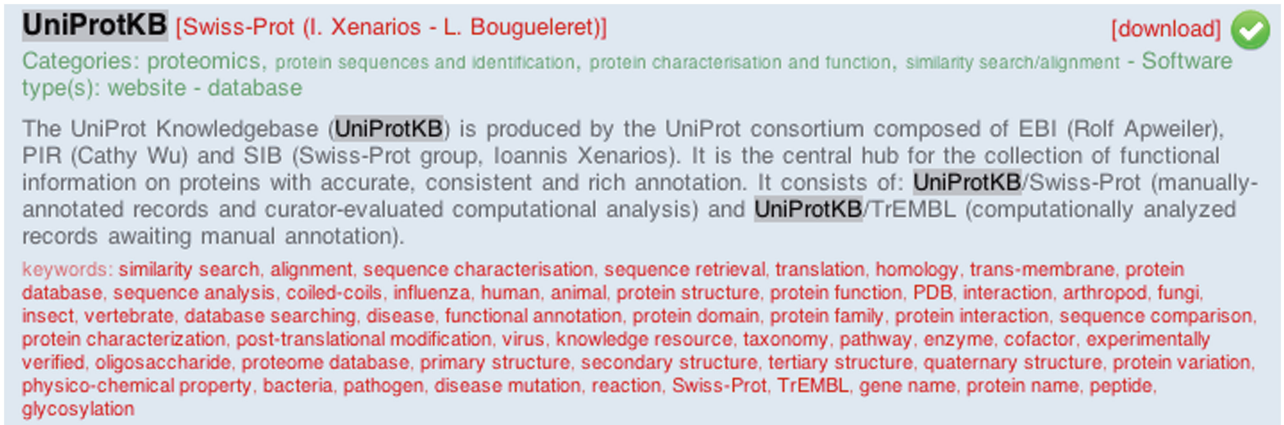

Figure 2. Detailed information describing a resource.

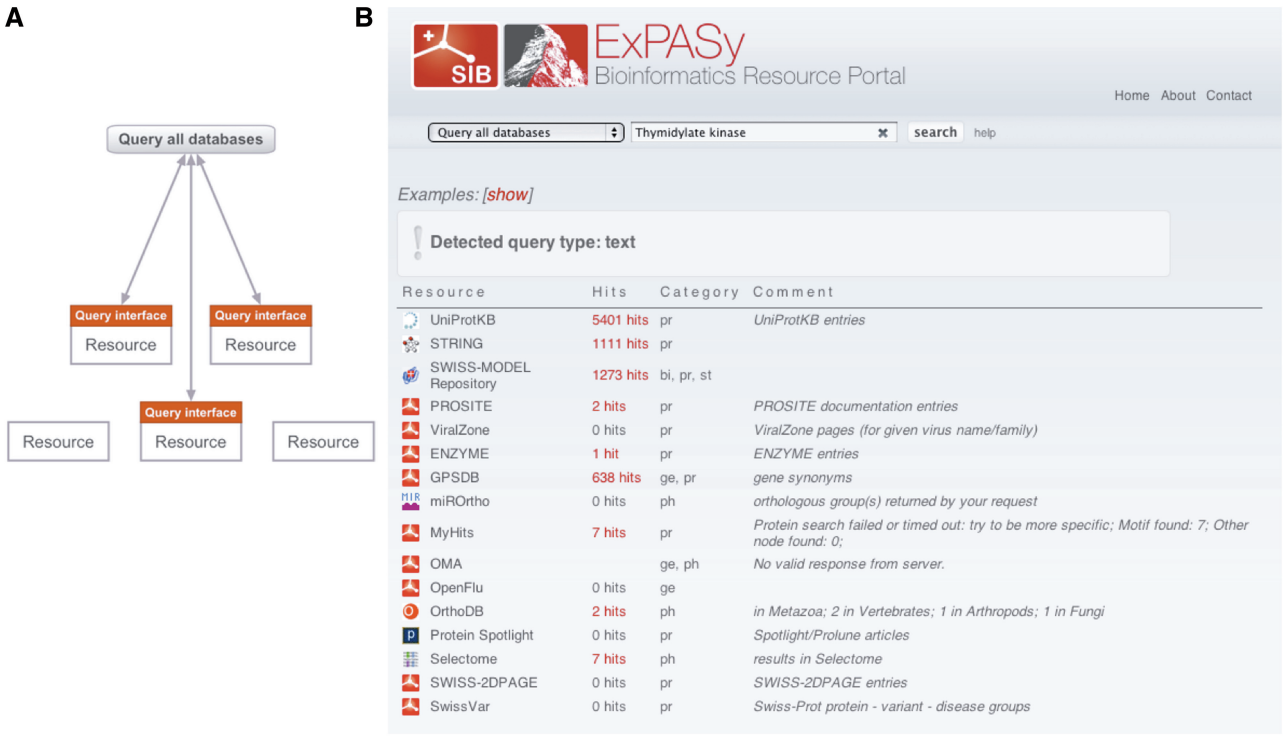

Figure 3. Overview of resources that can be queried via the cross-resource search interface (**A**) (Query all databases). If a resource implements the cross-resource search interface (**B**) (indicated by the red box named 'Query interface' and the arrow leading to the resource), ExPASy will query the resource directly. Other resources that do not implement this specific interface (even if they have other query features), are not included in ExPASy's parallel query.

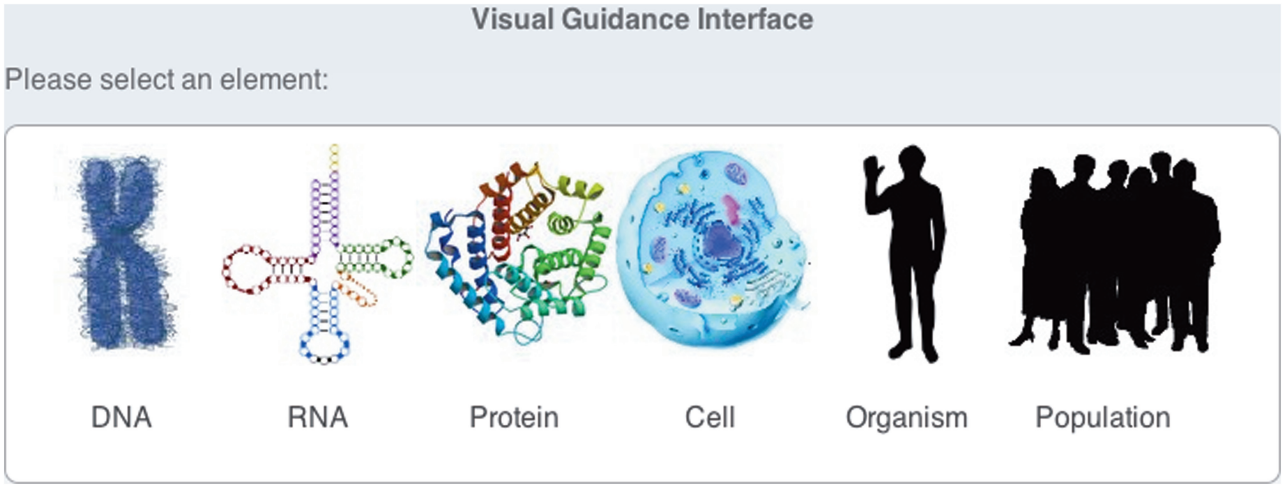

Figure 4. Entry point for the visual guidance interface—includes categories to be selected.

etc., and sends the query only to resources supporting the specified query type. The search can be extended beyond these automatically selected resources, by clicking on the button 'Search again as text', which initiates a full-text search to all registered resources.

Visual guidance

As an alternative to the standard, text-based interface that relies on scientific categories (which mainly targets expert users), resources can be accessed via a visually attractive interface. For instance, a biologist working at the cellular level might want to find corresponding tools and databases in another domain. The different graphical entry points are depicted in Figure 4. Once selected, databases and tools will be displayed (Figure 5).

Once a category has been selected, the list of resources shown can be narrowed down by a classical keyword-based search feature, or by a tag cloud showing the most frequent yet discriminative keywords among displayed resources. In the example in Figure 5, the

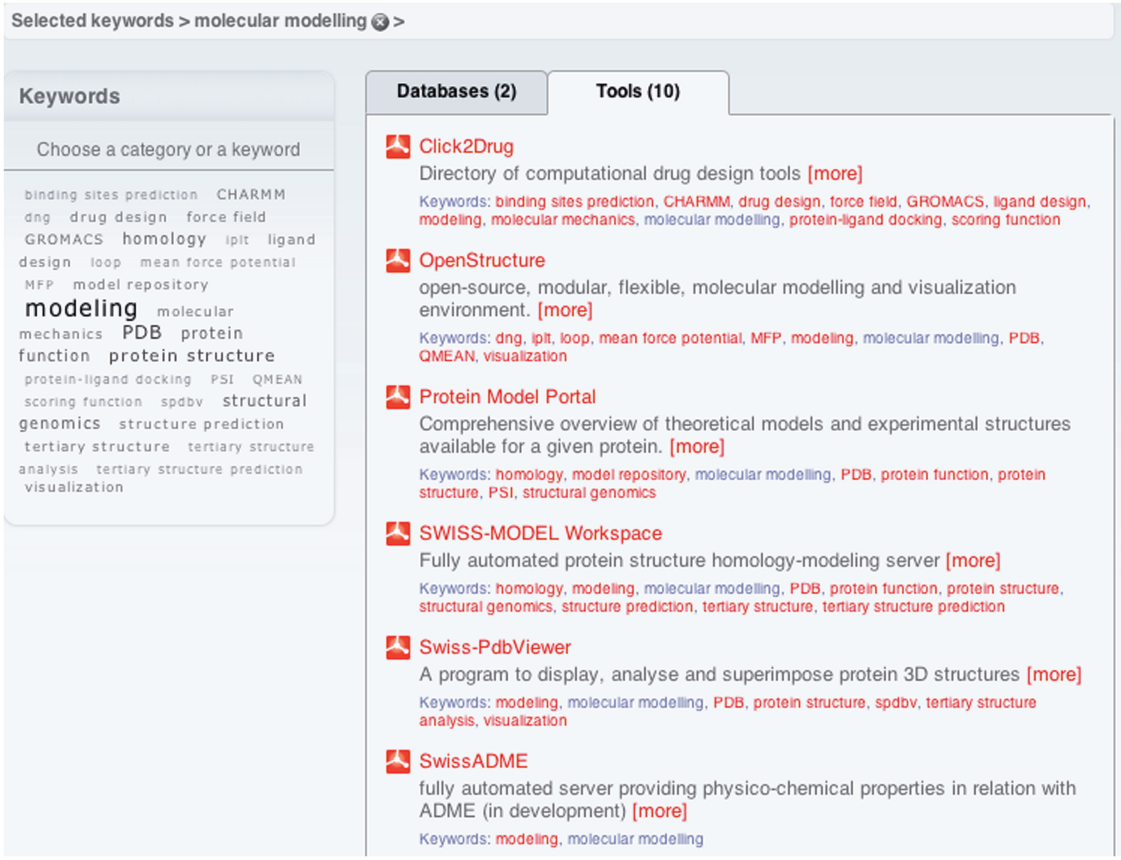

Figure 5. Tag cloud in the visual guidance interface.

keyword 'molecular modelling' has been selected (cf. 'Selected keywords' in the top left of Figure 5), and resources related to this keyword are indicated in the right pane.

Monitoring (availability and usage checks)

The federated portal lists resources from numerous groups in various locations with different levels of expertise. However, a common important goal is to provide high-quality resources that are available to the research community, i.e. service downtime should be reduced to a minimum. ExPASy provides a monitoring feature that periodically checks if a service is accessible and sends Email notifications in case of errors. The status of a resource is then displayed on the portal (green sign if ok, red sign if not ok). Additionally, the service up-time is displayed (Figure 6).

If a resource provides a web service interface, detailed functionality tests can be done. Since the EMBRACE registry (3) already provides a service to register and execute client-side tests for SOAP services, its interface is used and interrogated for service information. All these checks are performed by a Python script, which can also be extended to add other useful tests. For instance, one can test for additional functionality using example queries.

Service availability is one issue. Another question for users of a portal is: how popular is a specific resource? The information is not visible publicly (i.e. the statistics only are accessible to resource providers and portal operators) but the most popular resources on the portal are ranked according to internally collected usage statistics, e.g. UniProtKB appears on the top of the page.

Additional changes and hints for established users

We have also used the opportunity to critically review the resources on the previous ExPASy server, update them and phase out some of them. More than 40 resources that were used frequently and for which all source codes were still available, have been repackaged, recompiled and migrated to 64-bit architectures. Whenever needed, the user interfaces have been adapted. The repackaging also resulted in new URLs for most of the resources, i.e. the package name is now explicitly visible in the URL (either *.expasy.org/<resourcename> or <resourcename>.expasy.org). Example of a new URL: http://web.expasy.org/translate instead of http://www.expasy.org/tools/dna.html. However, old URLs will remain available for some time and are automatically redirected to new URLs.

The portal also has a new way to request help, i.e. we provide a single point of entry rather than several different Email addresses. Typically, we ask users to specify the resource for which they have a specific question, i.e. one can select from all existing resources. In this way, the

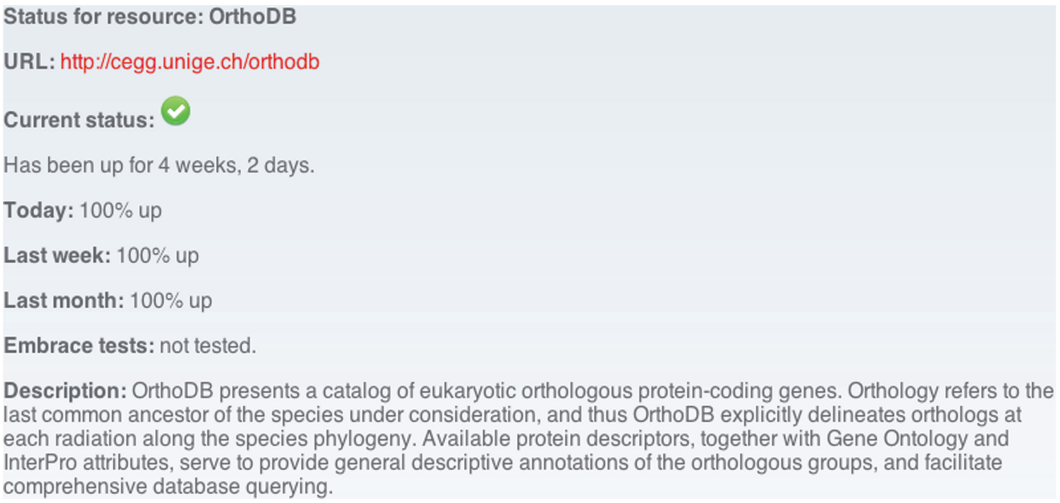

Figure 6. Monitoring information for a specific resource on the portal. This information is available if one clicks on the resources status sign displayed on the top right in Figure 2.

ExPASy helpdesk can better respond to the user's questions and address them to the correct people. Although resources are provided by several different groups, all providers can be contacted via a single form. That is a substantial improvement with respect to the previous version. Users only need to remember one place to ask questions (i.e., to post comments), whereas the ExPASy team takes care of finding the right contact person, i.e. we provide a professional helpdesk with first and second level support.

Mirror sites external to the SIB were discontinued for technical reasons but the new federated portal has a fail-over mirror at a second site in Switzerland (Basel) that can be used in case of technical interventions, etc.

The portal still provides access to the proteomics tools links page and its classifications via the menu item 'Links/Documentation – Proteomics software tools'. Additionally, the 'Life Science Directory' (previously known as Amos' links) is still accessible. However, both pages are static and are currently no longer maintained. This way, the portal still provides access to resources that are not maintained by the SIB. However, for a new release of ExPASy, it is planned that external resources (i.e. proteomics tools listed on the static page mentioned before) will be explicitly listed on the portal, included in the main search and displayed in a similar way as SIB resources. Additionally, some resources such as BLAST have been developed by external groups but the SIB operates the service and provides support.

CONCLUSION

Since 28 June 2011, the new ExPASy SIB Bioinformatics Resource portal has been online to serve the life sciences community. The new portal has been designed and developed by the SIB Web Team. It is fully operational and is accessed several thousand times a day by researchers from all over the world.

ACKNOWLEDGEMENTS

We thank the Steering Committee of the Web Team for the support: Ron Appel, Lydie Bougueleret, Fréderique Lisacek, Irene Perovsek, Torsten Schwede, Christian von Mering and Ioannis Xenarios. Thanks also to all ExPASy users who have provided feedback to improve the portal.

FUNDING

Swiss State Secretariat for Education and Research, in part. Funding for open access charge: SIB.

Conflict of interest statement. None declared.

REFERENCES

1. Appel,R.D., Bairoch,A. and Hochstrasser,D.F. (1994) A new generation of information retrieval tools for biologists: the example of the ExPASy WWW server. *Trends Biochem. Sci.*, **19**, 258–260.
2. Gasteiger,E., Gattiker,A., Hoogland,C., Ivanyi,I., Appel,R.D. and Bairoch,A. (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.*, **31**, 3784–3788.
3. Pettifer,S., Ison,J., Kalas,M., Thorne,D., McDermott,P., Jonassen,I., Liaquat,A., Fernández,J.M., Rodriguez,J.M., INB-Partners *et al.* (2010) The EMBRACE web service collection. *Nucleic Acids Res.*, **38**, W719–W723.

APPENDIX A1

In April 2012, the following SIB resources were listed on the portal:

AACompIdent, AACompSim, ALF, AllAll, Arlequin, Association Viewer, BayeScan, Bgee, Biochemical Pathways, BLAST, boxshade, ChIP-Seq, Click2Drug, CLIPZ, ClustalW, Codon Suite, COILS, Compute pI/MW, CT-CBN, Decrease redundancy, Dotlet, efmtool, ElMMo, EMBnet services, ENZYME, EPD, ESTscan, ExpressionView, FastEpistasis, fastsimcoal, FetchGWI /

tagger, FindMod, FindPept, Genome History, GlycanMass, GlycoMod, GlycoSuiteDB, GMM, GPSDB, HAMAP, HamapScan, HCD/CID spectra merger, ImageMaster / Melanie, ImmunoDB, ISA, Iscan, IsotopIdent, LALIGN, Linear Classification, MADAP, Make2D-DB II, MALDIPepQuant, MAMOT, MARA, MARCOIL, MassSearch, MIAPEGelDB, miROrtho, Mltree, MLTreeMap, MOSAIC Software Repository, Msight, MyDomains, MyHits, Myristoylator, Newick Utilities, neXtProt, nfswatch, OMA, OpenFlu, OpenStructure, OrthoDB, PANDITplus, PaxDb, PepPepSearch, PeptideCutter, PeptideMass, Phylogenetic Tree, Phylogibbs, pIcarver, Ping pong algorithm, PRATT, pROC, Prolune, PROSITE, Protein Model Portal, Protein Spotlight, ProtParam, ProtScale, QMEAN, QuickMod, RandSeq, RaxML, ScanProsite, Selectome, Sequence Similarity Maps (SSM), ShoRAH, SIBsim4, SIM, smirnaDB, Soaplab services, SSA, STRING, SugarBind, Sulfinator, SuperTree, SWISS-2DPAGE, SWISS-MODEL Repository, SWISS-MODEL Workspace, Swiss-PdbViewer, SwissDock, SwissParam, SwissRegulon, SwissSidechain, SwissVar, T-Coffee, TagIdent, TagScan, TCS, The Systems Biology Research Tool, TMPred, Translate, TreeGen, TriFLe, tromer, UniPathway, UniProtKB, UniProtKB/Swiss-Prot, Vertex Cover, ViralZone, Vital-IT, World-2DPAGE Constellation, World-2DPAGE Repository and ZFN-Site.